Quant Model

Zach Burns

April 30, 2019

```
#{r code = readLines('Code/graphs in paper.R'), echo = FALSE, message = FALSE, warning =
FALSE} #
```

Definitions

Assume the following general case of N task respondents responding to an open-ended prompt, each generating a different number of responses. Let $t \in \{1, 2, ..., N\}$ be a task respondent, each of whom produces j_t *jots*. Thus, $R_{t,j}$ is the j^^th jot produced by the t^^th task respondent. R is the set of all responses $\{r_{1,1}, r_{1,2}, ..., r_{1,j_1}, r_{2,1}, ..., r_{N,j_N}\}$. The count of this set we call the *total mass* of responses. We further define *frequency* as the number of times a jot appears in R. (Possibly can leave out the foolowing sentence) We use *jot* in contexts where the relevant level of analysis is the participant (e.g. a task respondent producing jots) and *response* in contexts where the relative level of analysis is the task (e.g. responses are scored relative to one another in the response set).

For example, imagine a task respondent is given 3 equally-sized LEGO bricks, A, B, and C that can be placed in positions 1, 2, and 3. An observer could track the positions of all bricks, and one possible jot is "BC.0.A" indicating that block B is stacked on top of block C in position 1, position 2 is empty, and block A is in position 3. In this task, there are 99 possible non-empty configurations (Garrett: I did calculate this!), and researchers might be interested in how unique each of the configurations are.

(In the example of a DT task, jots would be individual responses. More generally, jots can be any individuated state or response. In the example of a game with descrete game states, jots would be those states.)

We propose a measure of uniqueness called Mass-Adjusted Uniqueness Index (MAUI). MAUI is a response-level scaling index that defines uniqueness as a function of item and person parameters. (Garrett: I'm not sure exactly how to make this fit, but I'm sure we'll get after it at some point.) Specifically, MAUI defines uniqueness of a given response as the proportion of responses in the sample that were as- or less-frequently. MAUI has the following properties:

1) Within a given response set R, all responses that have the same frequency have equal MAUI scores
2) For a given response $r_{t,j}$ within response set R, all responses that have a lower frequency have lower MAUI scores

#These two need some exposition 3) The score is bounded between 0 and 1 4) The average of many MAUI distributions is, by definition, a uniform distribution

I actually think this next statement isn't true anymore, and probably isn't important anyhow

Finally, a response $r_{t,j}$ with frequency f will in general have a higher MAUI score in R_1 than R_2 if the total mass of R_1 is greater than the total mass of R_2. (though not guaranteed because it depends where that additional mass is in the distribution)

Suppose the following makeup of responses: Responses A_1, A_2, A_3, and A_4 were each given 5 times. Responses B_1, B_2, and B_3 were each given 4 times. Responses C_1, C_2, and C_3 were each given 3 times. Responses D_1 and D_2 were each given twice. Responses E_1, E_2, E_3, E_4, and E_5 were each given once.

The mass of responses A_n is therefore 20, B_n is 12, C_n is 9, D_n is 4, and E_n is 5, for a total response set mass of 50.

Suppose the context for this measure is a DT task. $R_{t,j}$ is a response to a prompt, perhaps it is "paperweight" in response to "think of all the possible uses for a brick." If paperweight is given by 25 participants, then the frequency of the jot "paperweight" is 25. 1) If the jot "doorstop" was also given by 25 participants, then "doorstop" and "paperweight" have the same score. 2) If the jot "build a house" was given by 30 participants, and "the saddest balloon ever" was given by 1 participant, then "build a house" has a lower score than "paperweight", which in turn has a lower score than "the saddest balloon ever." 3) If "paperweight" was given by 25 participants in response sets with total masses of 200 and 100, then it has a higher score in the former set.

Proof that weighted average of MAUI scores is always 1/2.

Let m_i be the frequency of the i$^{\text{th}}$ most frequently given answer.

Let $M_n = \sum_{i=1}^{n} m_i$; that is, the sum of all response mass frequencies from 1 (the largest) to n. If n is the rank of the least frequently given response (which will generally be those given by a single participant), then M_n is the total count of all jots given by all participants in the sample.

Therefore:

$$M_n^2 = (\sum_{i=1}^{n} m_i)^2$$

$$= (\sum_{i=1}^{n} m_i) * (\sum_{j=1}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i) * (\sum_{j=1}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i) * (m_i + \sum_{j \neq i}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i^2) + (\sum_{j \neq i}^{n} m_i m_j)$$

$$= (\sum_{i=1}^{n} m_i^2) + 2(\sum_{j<i}^{n} m_i m_j)$$

MAUI is defined as the mass of all responses given more frequently plus the center of the mass of responses given equally frequently. The weighted average of MAUI scores is therefore defined as:

$$\begin{aligned}
\overline{MAUI}_n &= \sum_{i=1}^{n}(\frac{m_i}{M_n})(\frac{\frac{m_i}{2} + \sum_{j=0}^{i-1} m_j}{M_n}) \\
&= \sum_{i=1}^{n}(\frac{m_i}{M_n})(\frac{m_i + 2\sum_{j=0}^{i-1} m_j}{2M_n}) \\
&= \frac{1}{2M_n^2}[\sum_{i=1}^{n}(m_i)(m_i + 2\sum_{j=0}^{i-1} m_j)] \\
&= \frac{1}{2M_n^2}[\sum_{i=1}^{n}({m_i}^2) + 2(\sum_{i=1}^{n}\sum_{j=0}^{i-1} m_i m_j)] \\
&= \frac{1}{2M_n^2}[\sum_{i=1}^{n}({m_i}^2) + 2(\sum_{j<i}^{n} m_i m_j)]
\end{aligned}$$

As shown above,

$$(\sum_{i=1}^{n} m_i^2) + 2(\sum_{j<i}^{n} m_i m_j) = {M_n}^2$$

The resulting quantity is 1/2 for all n. □